The people of the 24ᵗʰ annual Pacific Symposium on Biocomputing in Hawaii

This manuscript ([permalink](#)) was automatically generated from [dhimmel/psb-manuscript@c2f64b2](#) on January 4, 2019.

Authors

- **Daniel Himmelstein**
  ⓘ 0000-0002-3012-7446 · ⊙ dhimmel · 🐦 dhimmel

  Department of Systems Pharmacology & Translational Therapeutics, University of Pennsylvania

Abstract

Manubot is an open source tool for writing manuscripts on GitHub in markdown format. Manubot applies the git-based software workflow to scholarly writing, enabling enhanced transparency, collaboration, automation, and reproducibility.

This manuscript is the result of a *special working group* at the 2019 Pacific Symposium on Biocomputing that will introduce attendees to collaborative writing with Manubot. Each conference attendee is invited to write a small blurb on themselves and their research, by submitting a pull request to the manuscript repository at https://github.com/dhimmel/psb-manuscript.

The working group also covers how to write your next manuscript using Manubot and what features of Manubot can help biomedical researchers document and publish their computational research. For example, Manubot enables citation by persistent identifier to automate bibliographic metadata retrieval and formatting as well as allowing templating so results can be directly inserted from the analyses that produced them.

Methods

In this section, PSB 2019 attendees are asked to contribute a bit about themselves and their research. As part of the special working group, we thought this would be a helpful activity to introduce biocomputational scientists to writing with Manubot. For inspiration, here are some prompts:

- Introduce yourself briefly.
- What do you research? Include any relevant links to your lab or personal website.
- What is your favorite study from your career or what study was your biggest discovery?
- What was your first scholarly publication?
- Add a code snippet or mathematical equation.
- Add a figure with a caption. This could be a figure from your previous work or a picture of you in Hawaii.

Self-citations are explicitly encouraged, since one goal of this activity is to introduce attendees to the concept of citation by persistent identifier. By having attendees cite their existing works, we hope to show researchers how references can be created from just persistent identifiers, and how this can assist with collaborative and transparent authoring.

The markdown manuscript source has a section for each PSB 2019 attendee, generated from the online attendee list. Names are ordered alphabetically by last name. If you'd like to contribute, but are not already in the list, please insert your section at the appropriate alphabetical location.

For questions on how to use Manubot, see the usage documentation. More information on the tool and its inception is available in the project manuscript [1].

Attendees

Daniel Himmelstein

Greetings, I'm in the Greene Lab at the University of Pennsylvania and am the lead developer of the Manubot project. 2019 is my first PSB and I'm exciting to backpack around the Big Island following the conference.

My main area of research is integrating biomedical knowledge using hetnets [2,3]. However, I've also studied Sci-Hub, finding that it provides access to nearly all paywalled scholarly literature [4]. Perhaps my biggest discovery was observing an epidemiological association that higher elevation counties have lower rates of lung cancer, suggesting that oxygen is an inhaled carcinogen (Figure 1) [5,6].
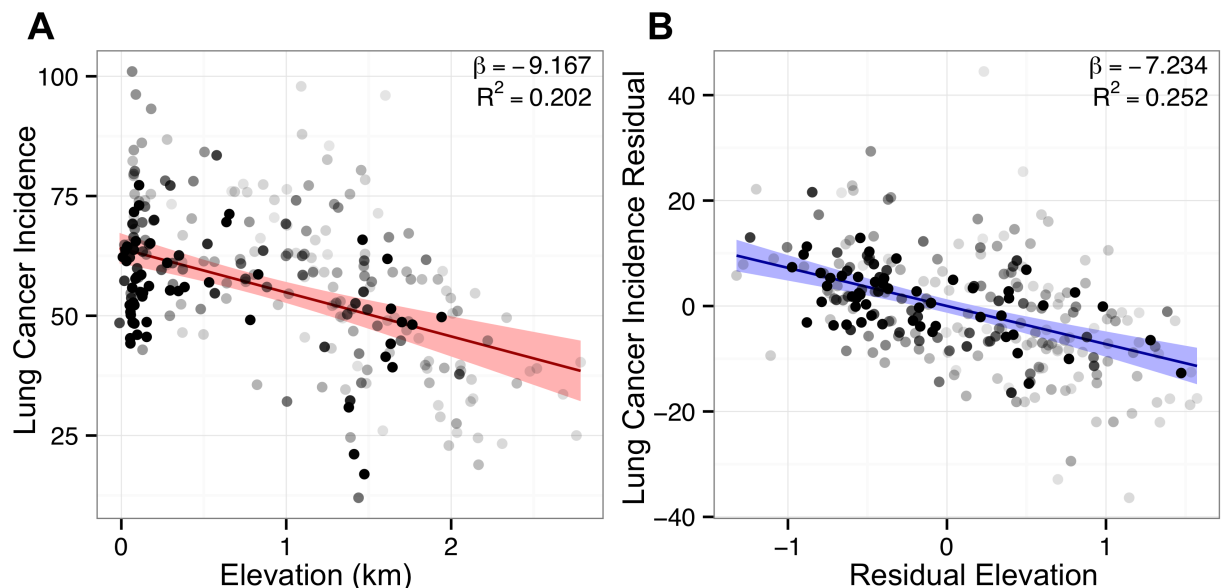

Figure 1: The association between elevation and lung cancer across Western U.S. counties.

I haven't done much text mining, but I did enjoy extracting attendee names for PSB from the online PDF. Converting the PDF to text in Python was as easy as:

```python
# https://stackoverflow.com/a/48673754
import tika.parser
parsed = tika.parser.from_file('attendees.pdf')
text = parsed["content"]
```

References

1. **Open collaborative writing with Manubot**
Daniel S. Himmelstein, David R. Slochower, Venkat S. Malladi, Casey S. Greene, Anthony Gitter
Manubot Preprint (2018-12-31) https://greenelab.github.io/meta-review/

2. **Heterogeneous Network Edge Prediction: A Data Integration Approach to Prioritize Disease-Associated Genes.**
Daniel S Himmelstein, Sergio E Baranzini
PLoS computational biology (2015-07-09) https://www.ncbi.nlm.nih.gov/pubmed/26158728
DOI: 10.1371/journal.pcbi.1004259 · PMID: 26158728 · PMCID: PMC4497619

3. **Systematic integration of biomedical knowledge prioritizes drugs for repurposing**
Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, Sergio E Baranzini
eLife (2017-09-22) https://doi.org/cdfk
DOI: 10.7554/elife.26726 · PMID: 28936969 · PMCID: PMC5640425

4. **Sci-Hub provides access to nearly all scholarly literature**
Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene
eLife (2018-03-01) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5832410/
DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410

5. **Lung cancer incidence decreases with elevation: evidence for oxygen as an inhaled carcinogen**
Kamen P. Simeonov, Daniel S. Himmelstein
PeerJ (2015-01-13) https://doi.org/98p
DOI: 10.7717/peerj.705 · PMID: 25648772 · PMCID: PMC4304851

6. **Unraveling the Ties of Altitude, Oxygen and Lung Cancer**
George Johnson
The New York Times (2016-01-25) https://www.nytimes.com/2016/01/26/science/unraveling-the-ties-of-altitude-oxygen-and-lung-cancer.html